Earning Inequalities Between and Within Nests

Mathieu Bunel, Jean-Pascal Guironnet

Centre de Recherche en Économie et Management
Center for Research in Economics and Management

University of Caen

University of Rennes 1

Working Paper

Earning Inequalities Between and Within Nests: A Multilevel Modeling Approach Applied to the Case of France

Matthieu Bunel

Jean-Pascal Guironnet

University of Caen Basse-Normandie, CREM-CNRS, UMR 6211

October 2011 - WP 2011-18

Earning Inequalities Between and Within Nests: A Multilevel Modeling Approach Applied to the Case of France

October 2011

First draft - please do not quote

Matthieu BUNEL[1] *and Jean-Pascal GUIRONNET*[2]

Abstract:

This paper presents a simultaneously study of the impact of gender and localization inequalities on the earnings of under-graduates. Using multilevel modeling, the framework draws both individual-level (i.e., pertaining to the individual elements of groups) and aggregate-level (i.e., pertaining to the group as a whole) data under a single specification, in order to study their potential interactions. These inequalities are studied with respect to young workers who left higher education in 2004 and who had a full-time job in the private sector three years after graduation (i.e., in 2007). To take into account the process of selection for employment, our multilevel model uses the Heckman two-step procedure. Following this approach, Occupational Groups (OG) are found to capture 59.4% of the earning heterogeneity whereas Employment Area (EA) nests capture 7.6%. This 59.4% figure is explained by two phenomena: (*i*) OG are dominated by seniors, and (*ii*) OG are dominated by males with higher earnings. These group characteristics also influence gender inequalities: there is a higher wage penalty for females in (*i*) OG dominated by males, and (*ii*) OG dominated by senior workers. In contrast to the gender gap, immigrant inequalities manifest closer links to EA. Policy implications are derived from our results.

JEL Classification: I23, R23, D31, J31

Keywords: multilevel models, wage differentials, gender inequality, local labor markets

[1]CREM (UMR 6211), Université de Caen Basse-Normandie, 19 rue Claude Bloch, 14000 Caen, E-mail: *mathieu.bunel@unicaen.fr.*

[2]CREM (UMR 6211), Université de Caen Basse-Normandie, 19 rue Claude Bloch, 14000 Caen, E-mail: *jean-pascal.guironnet@unicaen.fr.*

1 Introduction

Level of training, as defined by the number of years in education, is a major determinant of wage levels in starting a career (e.g., Spence, 1973 and 1974). Earnings are also influenced by other factors, such as the faculty in which training took place (social sciences, exact sciences, etc.), the candidate selectivity profile of the type of institution attented (Business School *vs.* Academic Faculty), as well as gender, ethnic origin, etc. Numerous studies have assessed the effects of such individual characteristics on earnings; however, the level of individual characteristics is marked by significant data heterogeneity. To decrease it, individuals can be grouped into homogenous social units, and these clustered factors can also be used to explain earning inequalities. On this approach, then, the individual characteristics of the worker are not the only explanation of earning gaps, but are supplemented by the characteristics of the group to which an individual belongs.

First, the earning variance is mainly influenced by occupation, since women and men have different choices in the range of jobs available to them. Studies have shown that gender discrimination restricts women's choice in their range of job: in particular, gender inequality increases gender discrimination through, a cultural 'devaluation' of work done by women (e.g. Tam, 1997). Occupational Groups (OG) dominated by females should thus manifest lower earnings (Bergmann, 1974; Datta Gupta, 1994; Simon, 2010). In addition, where an OG is dominated by women, this has unequal influence on males and females (Budig, 2002; Huffman, 2004). Accordingly, this framework proposes to use multilevel modeling to study the 'within-job' effects as well as 'between-jobs' effects on earning inequalities. One of the purposes of the paper is thus to answer the following question: Does the concentration of women inside an occupation influences the earning gap? Furthermore, and as a development on previous studies, this framework also explores the influence of mean age of an occupation upon the earnings gap: Does the age structure of an occupation influences the wage differential?

Second, for most workers, access to suitable employment is constrained by having to search for jobs in the local labor market rather than across the entire country (Van Ham et al., 2006). In France, the characteristics of the labor markets differ strongly from one area to another with respect to such things as the importance of manufacturing industries, unemployment, and wage policy. Using the French Labor Force survey, Minni and Vergnies (1994) show that the probability of finding a job is strongly influenced by the area of study, especially for women. This spatial mismatch explanation can be reinforced by the spatial mismatch hypothesis (Kain, 1968, 1992), and some papers try to specify the interaction between them such as, for example, the 'gendered' spatial mismatch hypothesis (Hanson and Pratt, 1995). Recent papers have applied this approach to the French case (for a survey, see Domingues Dos Santos et al., 2009).

Numerous studies have focused upon the impact of such characteristics to underline the existence of gender/origin inequalities or spatial segregation: however, most analyze these dimensions separately. By using multilevel modeling, the framework proposed in this paper draws these individual and aggregate dimensions together within the same model in order to analyze their potential interactions. This study thus aims to addressGoshen's (1991) question: "Is it who you are, what you do, or where you work" which influences the wage differential?

The multilevel model, then, allow us to deal with both individual and group characteristics. The 2006 French Census, the "Génération 2004" and, the DADS ("Déclaration Annuelles de Données Sociales") databases are used to estimate our applied measurement method. By providing a simultaneous estimation of the wage variance at individual and at the occupation/EA levels, our approach makes it possible to arrive at a comparative picture of their differential effects upon earnings. This type of study thus provide a new framework for inequality explanations, and provides new recommendations for policy-makers. As well as estimating the usual within-group effects, the introduction of between-groups estimations in line with the various clusters provides a new interpretation of individual opportunities within a context of a discriminatory labor market.

The main result of our study is that jobs associated with females manifest a cultural devaluation of such work, which increases the gender wage gap. This inequality is, however, lower in occupations dominated by young males.

The first section of this framework presents our data and discusses various descriptive statistics; the second section describes our choice of applied measurement method; and the last section comments on our econometric results before concluding.

2 The data

The main data source used in this paper is the "Génération 2004" survey carried out by the Centre for Studies and Research on Qualifications[3] in 2007. This survey draws on the "Generation 92" and "Generation 98" databases,[4] which are important surveys for the analysis of the initial years of education-to-work transition and occupational integration in France.

This survey is based on a sample of 33,655 young people who left school in 2004, representing a weighted number of 737,000 students. This database is representative of the school-leaver cohort of the French education system in 2004, and provides detailed information about the individuals' first three years of working life, occupational status, and other socioeconomic characteristics.

Students who have decided to pursue graduate studies, and those who do not have a bachelor's degree (first degree of higher education), have been dropped from the sample. For undergraduates, initial training is the main determinant of professional insertion, in contrast to the lower levels of educational attainment. Hence, students who have been trained in health faculties and in hospitals have also been dropped from the sample, since in France there are various *numerus clausus* for these careers. Lastly, individuals who have not stated their faculty of training, their location of work, or who work as civil servants, have been also dropped. Our study is thus based on a set of 12,215 individuals, which represents a weighted number of 283,073 French graduates of higher education. Among them, 9,805 are full-time employed, representing a weighted number of 224,521 workers in the private sector. Within this sample, 5.6% are still unemployed three years after leaving higher education. Despite the uncertain economic situation prior to the financial crisis at the end of 2007, this rate of unemployment seems to indicate an improvement in young

[3] Centre d'études et de recherches sur les qualifications (Céreq).

[4] For previous applications of this database, see Barros et al., (2011a, 2011b); and Guironnet and Peypoch (2007).

people's employment conditions in the French labor market.[5] This sub-sample thus yields individual information for the first level of analysis.

The two second levels of the analysis are OG and EA. The information concerning OG is drawn from an exhaustive administrative file called DADS (Déclarations Annuelles de Données Sociales – Annual Declarations of Social Data) for the year 2007. A sub-sample of 1/12th is available from the French institute of statistics. This file provides information for numerous OG such as the share of women, the training, the working time and the age of the worker in the private sector. The French Institute of Statistics defined more than 495 OG. We merged this file with our sub-sample from "Génération 2004"; some OG are whether not present in the individual database, and some have too few individuals (less than ten workers) to be useful. In such cases, groups were merged, leading to a final list of 171 OG.

For spatial information, we use the 2006 French Census[6] collected by the French Office of Statistics. The spatial unit used here is the EA. The Office of statistics lists 348 EA, defined by the fact that most of the people who live in the respective areas also work there. The French Census gives information on gender, origin, age, degree, and occupation of the individuals. It is feasible to calculate certain aggregate statistics about the local labor markets with respect to the EA, such as the unemployment rate and the unemployment for young people, and this file was merged with our sub-sample. Again, for some EA there are none or too few individuals; accordingly some were merged, resulting in a final list of 250 area clusters.

2.1 Individual-level variables

The "Generation" survey contains very accurate information on schooling (educational level, faculty, honors) in the first initial years of the school-to-work transition (tenure, experience, unemployment periods), and on the professional situation three years after graduation (wage, industry, type of contract). Socioeconomic information regarding the

[5] In accordance with the French Office of Statistics, the rate of unemployment of graduates four years after leaving higher education decreased in 2007.

[6] For more details on the new French Census, see Gobinot (2008).

individual is also available (sex, age, ethnic origin, marital status, number of children, parents' profession, and localization). This set of variables provides additional information which has generally been unobserved in previous studies: for example, honors can be considered as a proxy for student quality. Descriptive statistics are presented in **table #A1** in the appendix.

Furthermore, a part of "Génération 2004" is devoted to surveying individuals about their feelings of discrimination in job recruitment. Despite the subjectivity of the query and the potential interaction effects between questions, the answers provide initial information about labor market inequalities. Among them, the major inequalities are picked out by three criteria: ethnic origin, gender, and localization. Based on these criteria, 8.62% of the sample feel they have been discriminated against at least once in a job application. This result can be decomposed in line with the three main inequality factors (see table #1).

Table 1. Subjective Factors of Inequalities

Discrimination due to:	%	Share of Women	Share of Foreigners
Complete Sample (Weighted Number of 283,073 Graduates from Higher Education)			
Origin	4.65	43.96	72.49
Gender	4.08	87.55	22.06
Localisation	1.79	50.14	52.93
Sub-Sample of Full Time Workers (Weighted Number of 224,521 Workers)			
Origin	4.14	39.47	70.90
Gender	3.75	86.44	20.81
Localisation	1.59	47.74	49.52

Source: Génération 2004.

Inequalities are lower within the full-time workers sub-sample, implying an initial selection process which should be taken into account in our study on earnings. From table #1, the inequalities declared by individuals are due principally to: (i) gender, especially for women, and (ii) ethnic origins. The spatial segregation in job recruitment can be

Table 2. Objective Factors of Inequalities

Variables	Total (%)	Females (*diff.*)	Immigrants (*diff.*)
Job characteristics			
Full-time job	79.3	-16.9	-6.5
Temporary contract	27.1	5.89	6.97
Earnings	1607.3€	-11.1	1.7
Educational Courses			
Exact sciences	72.2	-24.2	-8.9
Under-graduate	34.3	-7.5	5.8
Graduate	36.3	3.9	-3.3
Master	29.5	3.69	-2.41
Academic	6.9	5.07	3.36

Source: Génération 2004.

also observed to have a potential interaction effect with ethnic origins: foreign workers probably live in areas marked by higher levels of discrimination.

Table #2 presents some objective statistics. The gross gender wage gap is around 12%. This rate is quite similar to those published by the OECD for all French full-time workers. This gap can partly be explained by three main observed differences between studies and choice of job by women and men: (*i*) women have lesser tenure and experience (around one month) than men; (*ii*) women are much less likely to have studied in a faculty of exact sciences; (*iii*) 30.5% of women in the labor market have a temporary contract, *versus* 24.6% for men. The main point that should decrease the wage gap, however, is that women have a higher level of education than men (38.5% are graduate and 31.5% have a master degree against 34.7% and 27.9% for men). Immigrants show similar trends but with a lower educational level: the results with respect to ethnic inequalities seem to be more mixed.

We can reasonably expect that these inequalities, and their corresponding interaction effects, will have an impact upon earnings. From table #2, the wage gap is clear for females but not obvious for immigrants.

2.2 Job-level variables

Recent studies have shown that the root of the gender wage gap can parlty be explained by reference to the workplace. For instance, several papers have pointed out that female-dominated jobs offer lower compensation than male-dominated ones (Datta Gupta, 1994; Datta Gupta and Rothstein, 2005). Simon (2010) used microdata from nine representative European countries to outline how female segregation into low-paying workplaces is an outstanding feature of the gender wage gap across the European economies. Furthermore, Ruijter and Huffman (2003) and Huffman (2004) observed in the Netherlands that male remuneration decreases when the percentage of women in an occupation increases.

To take both these effects into account, we followed Ruijter and Huffman (2003) in defining three dummy variables corresponding to the gender composition of each OG. These dummies reflect male-dominated, gender-mixed, and female-dominated professions. In our sample, approximately 36.2% of the workers in the OG, in which young people are employed are women. Male and female-dominated professions are defined as follows: all professions with more than 70% females are female-dominated, while all occupations with less than 20% female are male-dominated. The remaining occupations are defined as gender-mixed.

Age structure is another workplace characteristic that may influence individual wages. This link is, however, less documented than the previous one, and much more difficult to determine with accuracy. Research has not produced a clear and consistent pattern of results supporting the direction and the degree of such a link.

One approach is to find out whether firms with younger workers are more successful than those with older workers, and, if this is the case, are the workers receiving higher wages? Using French data, Aubert, Caroli, and Roger (2005) observed that in innovative firms the wage bill share of younger workers is higher than that of older workers.

A second approach draws on relational demography studies (Riordan, 2000). This research tries to determine whether more similar individuals - those forming a social unit in terms of age characteristics or otherwise - will be perceived more positively. For example, younger workers employed in senior dominated workplaces may be perceived to have lesser experience. Few papers identify such link, however: Ostroff and Atwater (2003) pointed out that managers in groups younger than age forty receive lower wage than managers in groups over than age forty.

To control for such an effect in our analysis, dummy variables are introduced to take the age composition of the OG into account. In our sample, approximately 38.6% of the graduates from higher education are in the OG dominated by workers under thirty-five years old. An OG is defined as senior-dominated when this percentage is less than 30%, as age-mixed when the percentage is between 30% and 50%, and as young-dominated otherwise.

Finally, dummy variables are also introduced to characterize occupations which are working-time intensive or less-working-time intensive. In our sample the average annual working time is 1,631 hours (approximately 35 hours per week). OG are defined as working-time-intensive when the average working time is higher than the third quartile (1,729 hours), and as less-working-time intensive when it is lower than the first quartile (1,573 hours).

3 Econometric Modeling

In order to study the determinants of wage inequalities upon the French labor market, we focus on individuals as nested by OG or as nested by EA. The influence of localization and OG on earnings is usually controlled by dummy variables. If we consider that the data are clustered in groups, the use of Ordinary Least Squares (OLS) to determine the influence of each group on the intercept and on the slope would require the introduction of a huge number of parameters (Bryk and Raudenbush, 1992; Goldstein, 2003). The introduction of numerous fixed effects in OLS regression can produce a misspecification: multilevel models can deal with this problem (Woodridge, 2002). The multilevel regression analysis is based on the estimation of two parameters (mean and variance) for the intercept and the slope analyzed in level 1. In what follows, we present our measurement method in more detail.

3.1 Unconditional Model

As a preliminary step, we estimate the simplest feasible model, called the unconditional model, which is formally equivalent to a one-way analysis of variance with random area effects:

$$Y^*_{ij} = \beta_{0j} + \epsilon_{ij} \tag{1}$$

where Y^*_{ij} is the logged monthly wage of person i in group j, and β_{0j} is the mean of the dependent variable. The intercept is expressed as a linear combination of a random deviation specific to a given group j with an intercept γ_{00}. In this case, when the intercept is area-specific, we get:

$$\beta_{0j} = \gamma_{00} + u_{0j} \tag{2}$$

The term u_{0j} corresponds to the residual error term at the area level. This residual term is assumed to have mean zero and to be independent from the residual errors ϵ_{ij}. This model decomposes the variance between the two independent components as follows:

$$Var(Y_{ij}) = Var(\gamma_{00} + u_{0j}) = \tau_{00} + \sigma_s^2 \tag{3}$$

where σ_s^2 is the within-area variance and τ_{00} is the between-area variance. The intra-class correlation coefficient in area (ρ_s) is given by

$$\rho_s = \tau_{00}/(\tau_{00} + \sigma_s^2). \tag{4}$$

3.2 Conditional Model

The next step introduces two types of covariates at the individual level and at the aggregated level. The individual's covariates reduce the intra-class correlation as all ordinary single level models. For instance, we might generalize (equation #1) as:

$$Y_{ij} = \beta_{0j} + \beta_{1j}\left(gender_{ij}\right) + \beta_2\left(X_{ij} - \overline{X}_{..}\right) + \epsilon_{ij} \tag{5}$$

with $gender_{ij}$ is a dummy variable coded 1 for women and 0 for men, and X_{ij} are covariates. We used a grand mean centering by subtracting the grand mean $\overline{X}_{..}$ from the corresponding individual score. By using this centering the intercept could be interpretable as the expected value of the outcome variable when the covariates have their mean value. This specification leads us to distinguish between-parts effects (representing the effects of the group OG or EA), and within-part effects (representing individual performance relative to peers (Kreft, et al. 1995; Hox, 2002)).

The introduction of aggregated covariates allows us to examine whether the level-2 explanatory variables explain the intra-class correlation.

$$\begin{aligned} \beta_{0j} \quad = \quad & \gamma_{00} + \gamma_{01} \left(gender\ composition\ dummies_j \right) \\ & + \gamma_{02} \left(age\ composition\ dummies_j \right) + \gamma_{02} \left(working\ time\ intensity\ dummies_j \right) + u_{0j} \end{aligned} \tag{6}$$

All these covariates are grand mean centering.

The last step introduces cross-level interactions. In the specification with level-2 corresponding to OG in our model, we can test whether the gender wage gap differs for occupations according to their gender composition (female-dominated, gender-mixed and male-dominated) and to their age composition (young-dominate, age-mixed and senior dominate). The introduction of the aggregate level covariates decreases the intra-class correlation.

$$\beta_{1j} = \gamma_{10} + \gamma_{11} \left(gender\ composition\ dummies_j \right) + \gamma_{12} \left(age\ composition_j \right) + u_{1j} \tag{7}$$

The parameters γ_{11} and γ_{12} show whether the effect of gender on wages is larger or smaller in female-dominated (or male-dominated) professions or in young-dominated (or senior-dominated) professions. u_{0j} and u_{1j} are level-2 random effects assumed to be uncorrelated and with mean zero.

3.3 Model with selection

The estimation is performed using the Heckman two-step method as follows. In a first step a probit model is used to estimate the probability that a young individual is in the labor market and full-time employed. A probit model is estimated to calculate an Inverse Mills Ratio (IMR). The second step estimates the equation (equation #5) augmented by the IMR. By including the IMR in our multilevel models and by estimating the model using the two steps strategy the true variance-covariance is not correctly estimated. Note that in a multilevel model with selection the exact formulation of variance-covariance is not easy to find. The bootstrap method provides a way to substitute the calculation of the asymptotic form of such distribution (Efron, 1979). Globally, the bootstrap method yields an approximation of the distribution which is more accurate than the approximation obtained from first-order asymptotic theory; the drawback is that it is time consuming. This time is, however, tractable with the use of a powerful processor. We use 1,000 bootstrap samples, which is computationally demanding but gives sufficient accuracy.

4 Results

In a first step, table #A2 presents the results of the probability of getting a full-time job, as estimated by the probit model. These results will be briefly discussed, since the aim of this paper requires that we draw heavily on the multilevel estimations. The estimation has produced expected signs for all variables.[7] Some positive correlations with the probability of finding a full-time job are found: (*i*) vocational courses; (*ii*) honors (which can been seen as a proxy for student quality); (*iii*) schooling level, especially for students in exact sciences; (*iv*) males, especially with children, contrary to females for which we see the opposite effect (Frederiksen, 2008); and (*v*) workers in Ile de France and Paris. Some negative correlations with the probability of finding a full-time job are also found: (*v*) parents with children, especially for women; (*vi*) individuals of foreign origin; (*vii*) months of unemployment in the first year after leaving education; and (*vii*) the area rate of unemployment (which can be seen as a proxy for labor market tightness). Following these results, the usual inequalities are found in getting a job. The most significant among them in terms of employment is gender discrimination.

[7] The Business School variable has been dropped from estimations since all students from these faculties have found a job.

In this econometric model, the following variables are estimated as valid instruments (not correlated with earnings): children and the composite variable formed by crossing the children and gender variables, which have no effect on earnings. The variables of area unemployment rate crossed with Paris and Île de France can also be considered as weak instruments. Concerning model quality, a better fit has been found for each multilevel model than the OLS regression with some aggregated dummy variables. Indeed, the multilevel model nested by EA represents 283 area dummies and for OG, we have 171 nests. Furthermore, the probit estimation allows us to calculate the usual IMR to take a potential selection bias (Heckman, 1979) into account.

The results of the multilevel analysis appear in tables #3 and #4. For the reliability of the text, the tables report estimated coefficients only for the main variables (complete estimations for the model Ic and IIb are available in the appendix). For the purposes of comparison, five models are estimated: the usual OLS regression; model Ia, which is a multilevel regression nested by OG; model Ib, which is a multilevel regression introducing level-1 covariates (degrees, faculties, temporary contract, honors, ERASMUS, parents' professions); model Ic, which introduces level-2 group control variables; and model Id, which is a complete estimation of both levels. Table #4 presents model IIa, which is a multilevel regression nested by EA whereas the model IIb includes only level-2 group control variables and model III is a complete estimation of the level-1 covariates and level-2 units. Models Ia and IIa express the variation of the logged wage as the estimated grand mean in the population (γ_{00}) more level-1 and level-2 random effects, in comparison to a one-way ANOVA. The variance of Y_{ij} is equal to $\tau_{00} + \sigma^2$, where $\tau_{00} = var(u_{0j})$, which yields the between-group variability, and $\sigma^2 = var(r_{ij})$ the within-group variability. R^2 usually gives the proportion of variance explained by the model. According to Bryk and Raudenbush (1992), in multilevel models the proportion of variance explained at level-1 is obtained by calculating $(\frac{\sigma_1^2 - \sigma_2^2}{\sigma_1^2})$ with σ_1^2 as the residual variance of the model Ia and σ_2^2 the residual variance of the comparison model. In model Ic, individual characteristics explain 11% of the earning variance, and by including all the level-1 variables this percentage grows to 12.5% (model Id). The level-2 units better explain the earning variance, with 72.2% for model Id.

Table 3. MCO and OG Multilevel Models

	MCO	Model Ia	Model Ib	Model Ic	Model Id
Between OG Variance (τ_{00})	-	0.036***	0.014***	0.019***	0.008***
Between OG Variance (τ_{01})	-	-	-	-	0.002***
Within-OG Variance (σ^2)	-	0.053***	0.053***	0.047***	0.046***
Intra-class Correlation (ρ)	-	40,6%	20,6%	28,7%	17,0%
Level-2 : R^2	-	-	61,9%	47,7%	72.2%
Level-1 : R^2	42.1%	-	-	11,1%	12,5%
AIC	-	-405.2	-533.9	-1524.0	-1683.8
IMR	0.084***	-0.231***	-0.220***	0.034***	0.038***
Intercept (β_0)	6.870***				
Intercept (γ_{00})		7.356***	7.127***	7.300***	7.158***
Female Dominate	-		0.183***		0.093**
Gender Mixed	-		0.199***		0.111***
Senior Dominate	-		0.138***		0.115***
Mixed Age	-		0.065**		0.054**
Intensive Working	-		0.053*		0.038*
No Intensive Working	-		-0.053*		-0.044**
PFT_g	-		1.700***		1.060***
PFT_g *Female Dominate	-		-2.252***		-1.277***
PFT_g *Gender Mixed	-		-1.362***		-0.714***
Female (β_1)	-0.114***				
Intercept (γ_{10})	-			-0.087***	-0.072***
Female Dominate	-				0.042*
Gender Mixed	-				-0.007ns
Senior Dominate	-				-0.057***
Mixed Age	-				-0.055***
Controls included	Level 1	None	Level 2	Level 1	All

** significant to 10%, ** significant to 5%, ***significant to 1%.*
Source: Génération 2004, Census 2006 and DADS 2007.

Table 4. EA and Two Hierarchical Level Models

	Model IIa	Model IIb	Model III
Between EA Variance (τ_{10})	0.006***	0.002***	0.002***
Between OG Variance (τ_{00})	-		0.008***
Between OG Variance (τ_{01})	-		0.002***
Within-EA Variance (σ^2)	0.076***	0.055***	0.045***
Intra-class Correlation (ρ)	7,2%	3,7%	19,6%
Level-2 : R^2	-	66.7%	63.2%
Level-1 : R^2	-	27.6%	11.7%
AIC	2871.9	-214.0	-1785.5
IMR	-0.446***	0.083***	0.040***
Intercept	7.3589***	7.279***	7.156***
Female Dominate			0.092**
Gender Mixed			0.110***
Senior Dominate			0.114***
Mixed Age			0.054**
Intensive Working			0.037*
No Intensive Working			-0.042**
PFT_g			1.052***
PFT_g *Female dominate			-1.285***
PFT_g *Gender mixed			-0.716**
Female			
Intercept		-0.114***	-0.071***
Female Dominate			0.043*
Gender Mixed			-0.008ns
Senior Dominate			-0.055***
Mixed Age			-0.055***
Controls included	None	Level 1	All

** significant to 10%, ** significant to 5%, ***significant to 1%.*
Source: Génération 2004, Census 2006 and DADS 2007.

For immigrants, the usual selection bias is found: migrants holding a job should be better than others since they have been discriminated against by the French labor market (see probit results from table #A2). Over the sample for employment, the mean wage of workers with French origins is around 1,598 Euros, whereas immigrants in full-time work have a mean wage of 1,644 Euros. Immigrants, however, have lower education and a higher probability of having a part-time job.[8] Therefore, workers from foreign origins those who have avoided first-job discrimination at recruitment probably have better

[8] Similar statistical results have been found and, in addition, ethnic inequalities seem to be decreasing in France (Céreq, 2008).

unobservable characteristics. IMR is significant in all models, testifying to the existence of selection bias as discussed before. When a gender variable is not introduced in the model estimations, IMR is positive, which implies that unobservable characteristics positively influence earnings. However, when the negative effect of the gender variable is not included (becoming unobservable), IMR is negative, implying that unobservable characteristics negatively influence earnings.

In table #A3 in the appendix, the estimated coefficients of model Id and IIc are quite similar but significantly different from the OLS regression. These multilevel estimations therefore show a significant heteroscedasticity bias in the data set. A solution is thus multilevel modeling to capture a part of the heterogeneity by some nests (which can be seen in the γ_{00} term). In the multilevel models, our estimated coefficients thus represent a mean effect whatever the nests which is not biased by group heterogeneities, in comparison to the OLS coefficients. With a grand mean centering, the intercept can be interpreted as the wage mean of a standard worker i.e., with characteristics (or variables) which take the value of the mean of the sample (when all independent variables are equal to 0). Only the tenure variable has a more stable estimated coefficient in all models. Tenure is thus a homogenous variable whatever the OG or the worker localization.

The intraclass correlation coefficient (ρ) defined by equation #4 gives the proportion of the variance in the outcome variable that exists between the level-1 units. From model Ia, ρ suggests that approximately 40.6% of the total earning variability is due to the differences across OG and 59.4% is attributable to the differences across individuals without covariates. Contrary to Groshen's (1991) results, differences between individuals explain the major part of the earning gap; differences between occupations, however, also explain a significant part of the earning gap. Our next estimations (model Ib) will try to explain the earning variability due to the differences between OG characteristics. When level 2 units are introduced, the unexplained variance decreases to 40% at 20%. This improvement is due to three group variables which are significant.

First, earnings are indeed positively correlated to the age of the SPC group, which appears to be as a positive outcome externality for a worker inside an OG composed of senior workers. Second, a negative correlation of earnings with the number of females within the OG can be seen[9] (Datta Gupta and Rothstein, 2005). From a statistical viewpoint, all occupations dominated by females have lesser wages than the sample mean (except in commercial jobs). For example, the role of secretary is dominated by females at 95.8%, and these women earn less than men in the building industry. This earning penalty for OG dominated by females is probably subject to a "cultural" devaluation of their jobs (Tam, 1997). PFT_g (probability og getting a Full-Time job in the OG[10]) captures the selection bias specific to the nests. PFT_g is then highly significant in model Ib: some group characteristics which produce a social unit in the OG are positively correlated to the earnings. This effect is, however, stronger for the social characteristics of OG dominated by males compared to those dominated by females which confirms our previous comments: the characteristics of being a male constituing a social unit positively influence the probability of being in a more profitable OG. If the level-1 units explain around 20 percentage points of the earning variance due to OG groups, 20.6% is not yet explained.

In introducing both levels, multilevel model Id explains around 11 percentage points of the earning variance due to the OG (τ_{01}). From model Ic, the gender coefficient shows a lesser bias, since a part of the heterogeneity of gender inequalities is captured by OG. Furthermore, the crossing variables between gender variable and group characteristics (i.e. level-2 units) can decompose the effect of gender inequalities (model Id). A part of these inequalities comes therefore from females who are less discriminated against in the OG with a high share of females. Conversely, they probably have more difficulties of getting a job in the OG dominated by males, but also have an earning penalty in these nests, in comparison to the males of the OG. OG dominated by males then do not have the same effects following the gender (Budig, 2002). In addition, females face lower levels of discrimination in OG with a high proportion of young men (where there are more than 50% in the OG): a part of the negative effect of gender is thus due to higher discrimination in some OG dominated by seniors.[11].

[9] See estimated coefficient of female and gender variable with their interaction with PFT_g variable.

[10] To calculate this term, we take the grand mean of the following operation: $\widehat{PFT} - \widehat{PFT}_c = \widehat{PFT}_g$, where $\widehat{PFT}$ is the probability of being employed as calculated by our probit, and $\widehat{PFT}_c$ is the group mean $\widehat{PFT}$.

[11] A similar decomposition has been tested for migrants; this decomposition is, however, not significant.

Table #4 presents the expected earnings and the associated gender inequalities for females as calculated from the estimated coefficients from model Id:

Table 5. Expected earnings (in log) by OG

	Female-dominate	Gender Mixed	Male-dominate
Young dominate	7,2079	7,2491	7,2166
Mixed age	7,2328	7,2451	7,2415
Senior dominate	7,2887	7,3299	7,2974
Earning Gap for Females (%)			
Young dominate	1,23%	-11,24%	-3,67%
Mixed age	-4,23%	-12,19%	-8,86%
Senior dominate	-4,75%	-10,11%	-7,54%

Source: calculated from model Id estimations.

Following the results from table #4, higher wages for women are expected in gender-mixed OG dominated by seniors. However, if females want to favor their relative earnings in comparison to males, the best situation to be in is OG dominated by females and young workers, although in this case there is a perceptibly the lower wage. The earnings of immigrants are not influenced by OG characteristics. However, some inter-group effects are probably significant, indicating that a part of the inequalities according to origin is captured by OG (see table A3 in the appendix).

From model IIa, EA nests significantly explain in comparison to the model Ia only 7.2% of the total variability of the earnings. If the area nests provide lesser the earning variance, their between-part effects are nevertheless significant. Level-1 covariates are included in model IIb. The positive effect of Paris and IDF variables seem to be overestimated in the usual regression. If the capital city captures the major part of the earning variance, some other spatial nests improve the "goodness of the fit" of the model.

In contrast to the OG, ethnic origin seems to show a higher degree of influence by EA, since this effect decreases and it is now not significant (not presented in the table). This result is confirmed by the previous statistics regarding individual feelings about inequalities: a high proportion of foreign workers discriminated by their origins seem also to be discriminated by localization (see table #1). By contrast, gender inequality seems to be more stable in model IIb.[12] Whereas females have the same return whatever EA, immigrants have a higher earnings variance on the basis of EA.

The final specification, model III, is a cross-classified model, i.e., a multilevel model with three levels: individual characteristics, EA, and OG. However, OG and OH are not perfect hierarchical clusters. In this case, estimations are based upon cross-classified data (Goldstein, 2003). In comparison to model IIb, the explained variance levels-1 and -2 decreases in model III. This result is probably due to a correlation between both variable levels and the OG and EA nests. The estimated coefficients are then not biased by this correlation. This estimation confirms our previous expectation with the ethnic origin variable being not significant. The incidence of ethnic origin is then overestimated in the usual OLS estimation, since this approach includes the inequality heterogeneity due mainly to the EA, which is not fully captured by area dummies.

5 Conclusion

The purpose of this paper is to study earning inequalities between and within social units. By using multilevel models, this framework improves first the "goodness of fit" of the usual regression, capturing some of the heterogeneity due to social unit clusters in the data set. The two-step econometric modeling also improves the quality of the multilevel model with the inclusion of IMR.

Following the first step, gender inequalities seem to be the main explanatory factor regarding inequalities in getting a full-time job. Further, ethnic origin is also significant. In view of the feelings expressed by individuals, it seems that these two factors represent the major causes of inequality: individuals, however, feel that the ethnic origin is the most pregnant factor; in the light of our probit estimation, this feeling seems to be overestimated.

These inequalities obviously have an impact on earnings. Our econometric modeling shows that the effects of ethnic origin and gender could have numerous and varied explanations. Following this approach, OG strongly improve the model quality, whereas spatial

[12] The level-1 units have been tested for EA clusters but they are not significant.

nests have a lesser but still significant effect. In investigating within- and between-part effects, all variables (except tenure) are heterogeneous between EA and OG.

From our multilevel estimations with OG, the data heterogeneity can be decomposed into 40.6% due to the difference across individuals, and 59.4% attributable to the difference across OG. The heterogeneity due to the difference between OG can be explained by the following effects: workers in a senior OG have higher incomes, as do workers in OG with a larger share of males. The first point can be explained by higher synergy effects inside a group with higher experience; the second can be explained by the cultural "devaluation" of jobs done by women.

These externalities are particularly worrying for females. The inclusion of group control variables show that women have lower wages in OG with a high proportion of males, and in OG with a high proportion of senior workers. The first of these facts indicates a double penalization, since women probably face more difficulties in coming to work in such OG in the first place, and also have lower earnings, compared to males. The second fact is probably due to a generational effect, implying that this cultural "devaluation" should be lesser for women in the labor market in the future.

In the last model, EA nests explain 7.6% of the data heterogeneity. The capital city captures the major part of area heterogeneity but some other spatial nests are also significant. In opposition to the gender inequalities, the ethnicity variable seems to be more influenced by EA. Individuals should adopt different strategies to avoid inequalities which are consequent on their ethnic origins and gender. There are certain OG in which females should work if they wish to avoid inequalities, and others if they wish to maximize their earnings, whereas migrants should take spatial segregation into account.

From the viewpoint of policy makers, attention should be focused on the inequalities due to ethnic origins. Due to the high correlation between origins and localization, a feasible recommendation would be to better integrate foreign individuals in less segregated areas. Concerning gender, this phenomenon seems to be decreasing, and we perhaps see evidence of a cultural shift with younger generations changing their mentality.

6 Acknowledgements

The authors are grateful for the useful remarks during the thirteen journeys upon longitudinal data in the analyses of labour market entitled "Les nouvelles ségrégation scolaires et professionnelles" in Toulouse (2011). The authors are also grateful for the helpful comments in the seminar of GATE in Saint-Etienne (2011).

7 References

Altonji J.G., Blank R.M. (1999), "Race and Gender in the labor market", in Ashenfelter, O. Et Card, D. (eds.), Handbook of Labor Economics, 3, Elsevier Science.

AubertP., E. Caroli and M. Roger (2005), "New technologies, workplace organisation and the age structure of the workforce: Firm-level evidence", *PSE Working Papers*.

Barros, C.P., Guironnet, J-P. and Peypoch, N. (2011a), "How to Quickly Get a Job? The Transition from Higher Education to French Labour Market by a Survival Model", *Applied Economics,* 43, 439-48.

Barros, C.P., Guironnet, J-P. and Peypoch, N. (2011b), "Productivity Growth and Biased Technical Change in French Higher Education", *Economic Modelling,* 24, 398-410.

Bergmann, B. (1974), "Occupational Segregation, Wages, and Profits when Employers discriminate", *Eastern Economic Journal,* 1, 103-10.

Budig, M. J. (2002), "Male Advantage and the Gender Composition of Jobs: Who Rides the Glass Escalator?", *Social Problems,* 49, 258-77.

Bryk and Raudenbush, (1992), Hierarchical linear models, Newbury Park, CA : Sage.

Cardoso A.R., (2000), "Wage differentials across firms: An application of multilevel modelling", *Journal of applied Econometrics*, 15, 343-54.

Domingues Dos Santos M., Y. L'Horty and E. Tovar, (2009), "Ségrégation urbaine et accès à l'emploi : une introduction", document travail CEE, novembre, 123

Efron, (1979), "Bootstrap Methods: Another Look at the Jackknife", *Annals of statistics*, 7(1), p. 1-26.

Goldstein H., (2003), Multilevel statistical models, Arnold, London.

Datta Gupta, N. (1994), "A specification test of the determinants of male-female wage occupational differentials", *Economic Letters*, 44.

Datta Gupta, N. and D. Rothstein (2005),The Impact of Worker and Establishmentlevel Characteristics on Male–Female Wage Differentials: Evidence from Danish Matched Employee–Employer Data", *Labour*, 19 (1) p. 1–34.

Frederiksen, A. (2008), "Gender Differences in Job Separation Rates and Employment Stability: New Evidence from Employer-Employee Data", *Labour Economics,* 15, 915-37.

Groshen, E. L. (1991), "The Structure of the Female/Male Wage Differential: Is It Who You Are, What You Do, or Where You Work?", *Journal of Human Resources,* 26, 457-72.

Guironnet, J-P. and Peypoch, N. (2007), "Human Capital Allocation and Overeducation: A Measure of French Productivity", *Economic Modelling*, 24, 398-410.

Hanson S. and Pratt G. (1991), "Job Search and the Occupational Segregation of Women", *Annals of the Association of American Geographers*, 81, 229-53.

Heckman, J.J. (1979), "Sample selection bias as a specification error", *Econometrica,* 47, p. 153-61.

Hedeker D., R.D. Gibbons et C. Waternaux, (1999), "Sample size estimation for longitudinal designs with attrition. Comparing time-related contrasts between two groups", *Journal of Educational and Behavioral Statistics*, 24, 70-93.

Hox J., (2004), *Multilevel Analysis: Techniques and applications,* Psychology Press, New-York.

Huffman, M., and Velasco, S. (1997), "When more is less", *Work and Occupations*, 24, p. 214–244.

Huffman, M., (2004), "Gender Inequality Across Local Wage Hierarchies, *Work and Occupations*, 31, p. 323–344

Kain J, (1968), "Housing Segregation, Negro Employment, and Metropolitan Decentralization", *Quarterly Journal of Economics*, 82, p. 175-197.

Kain J, (1992), “The Spatial Mismatch Hypothesis: Three Decades Later", *Housing Policy Debate*, 3, 371-460.

Kreft, I.G.G., de Leeuw, J. and Aiken, L. (1995), The Effect of different forms of centering in hierarchical linear models. Multivariate Behavioral Research, 30, 4-22.

Lainé, F. and Mahrez, O. (2005), "Jeunes de parents immigrés: de l'école au métier", *Travail et*

Emploi, 103.

Minni C. and J.F. Vergnies, (1994), "La diversité des facteurs de l'insertion professionnelle", *Economie et statistique*, 77-278, 45-61.

Ostroff C. and Atwater L. (2003), "Does whom you work with matter? Effects of referent group gender and age composition on managers' compensation", *Journal of applied psycholog*y 2003, 88(4), p. 725-740.

Rathelot R. et P. Sillard, (2009), "Zones Franches Urbaines : quels effets sur l'emploi salarié et les créations d'établissements ?", *Economie et statistique*, 415-416, 81-96.

Recchia A. (2010), "R-Squared Measures for Two-Level Hierarchical Linear Models Using SAS", Journal of Statistical Software, 32..

Riordan M. (2000), "Relational demography within groups: Past developments, contradictions, and new directions", *Research in Personnel and Human Resources Management*, 19,p.131-173.

Roy J. and X. Lin, (2002), "Analysis of multivariate Longitudinal Outcomes with nonignorable dropouts and Missing Covariates: Changes in Methadone Treatment Practices", *Journal of the American Statistical Association*, 97(457), 40-52.

Ruijter J. and M. Huffman (2003),"Gender composition effects in the Netherlands: a multilevel analysis of occupation inequality", Social Science Research, 32, p. 312-334.

Simon (2010),"International Differences in Wage Inequality: A New Glance with European Matched Employer-Employee Data", British Journal of Industrial Relations,48(2), p. 310-346,

Spence, M. (1973), "Job Market Signalling", Quarterly Journal of Economics, 87, 355-74.

Spence, M. (1974), *Market Signalling; Informational Transfer in Hiring and Related Screening Processes,* Harvard University Press, Cambridge.

Tam, T. (1997), "Sex Segregation and Occupational Gender inequality in the United States: Devaluation or specialyzed training?", *American Journal of Sociology,* 102, 1652-92.

Van Ham, M., Mulder, C.H. and Hooimeijer, P. (2001), "Spatial flexibility in job mobility: macrolevel opportunities and microlevel restrictions", *Environment and Planning* A, 33,921-40.

Wooldridge, J.M. (2002), *Introductory Econometrics: A Modern Approach,* South-Western College Pub (2eds).

8 Appendix

Table A1. Descriptive Statistics on the Student Weighted Numbers of the Samples

Variables	Description	All		Full-time only	
		Mean	std	Mean	std
Earning	dependent variable in Euros	1501,35	634,50	1607,34	613,70
Employed in full-time	binary variable	79%	0,41	-	-
Individual Variables					
Female	binary variable	48,1%	0,50	42,8%	0,49
Child	binary variable	7,7%	0,27	6,5%	0,25
Immigrants	binary variable	9,6%	0,29	8,9%	0,28
Executive father	binary variable	35,2%	0,48	36,2%	0,48
Executive mother	binary variable	18,3%	0,39	18,8%	0,39
Professional Transition					
Experience	coded in months	28,01	8,29	29,10	6,83
Tenure	coded in months	21,81	11,74	21,84	11,69
Unemployment	unemployed between 2004/05 coded in months	1,82	2,94	1,69	2,72
Studies Variables					
Bachelor	binary variable	38,4%	0,60	34,2%	0,55
Licence	binary variable	35,0%	0,48	36,3%	0,48
Master	binary variable	26,7%	0,44	29,4%	0,46
Business school	binary variable	1,6%	0,12	1,7%	0,13
Academic	academic training, binary variable	8,2%	0,27	6,9%	0,25
Honors	if the bachelor has been rewarded by honors	28,0%	0,45	30,3%	0,46
Erasmus	if the student has left at least six months in foreign country	8,4%	0,28	9,5%	0,29
Exact Science	binary variable	66,2%	0,47	72,2%	0,45
Localisation					
Paris	binary variable	7,0%	0,26	7,7%	0,27
IDF	île de France area without Paris	16,1%	0,37	17,8%	0,38
Others localization	binary variable	76,9%	0,45	74,5%	0,56
Unemployment area	rate of unemployment in %	19,9%	0,05	19,7%	0,05
Job Characteristics					
Manufacturing	manufacturing industries	27,5%	0,45	33,0%	0,47
Commercial	wholesale and retail trade	17,4%	0,38	17,6%	0,38
Business services	finance, insurance, and real estate and business services	35,1%	0,42	33,1%	0,38
Others services	personal services, entertainment, public services	20,0%	0,40	16,4%	0,37
Temporary contract	job contract, binary variable	26,5%	0,44	27,1%	0,44
Senior dominate	coded 1 if SPC has less than 30% of young workers	26,5%	0,44	23,9%	0,43
Mixed age	coded 1 if SPC has higher than 50% of young workers	60,5%	0,49	60,0%	0,49
Young dominate	coded 1 if SPC has more than 50% of young workers	13,1%	0,34	16,2%	0,37
Female dominate	coded 1 if SPC has less than 30% of male	16,4%	0,37	16,4%	0,37
Gender mixed	coded 1 if SPC has between 30% and 50% of male	17,9%	0,38	16,4%	0,37
Male dominate	coded 1 if SPC has more than 50% of male	39,6%	0,49	45,5%	0,50
Intensive working	work more than 35 hours	19,7%	0,40	22,8%	0,42
No intensive working	work less than 35 hours	22,8%	0,42	21,9%	0,41

Table A2. Probability of Full-Time Job

Variables	Probit
N	12215
Log-likelihood	-5242
AIC	10517
Intercept	0.555***
Licence	0.605***
Master	1.044***
Exact Science	0.845***
Academic	-0.694***
Honors	0.232***
Professional transition	
Log(Unemployment)	-0.162***
Log(Unemployment)*Paris	0.382***
Log(Unemployment)*IDF	0.259***
Family and origin	
Male	0.801***
Child	-0.744***
Child*Male	0.899***
Immigrants	-0.343***
Localisation	
Paris	0.385***
IDF	0.809***
Unemployment Area	-2.318***

** significant to 10%, ** significant to 5%, ***significant to 1%.*
Source: Génération 2004, Census 2006 and DADS 2007.

Table A3. Complete Estimations from MCO and Multilevel Models

Variables	MCO	Model Id	Model III
Studies characteristics			
Licence	0.092***	0.043***	0.045***
Master	0.314**	0.148***	0.150***
Exact science	0.053***	0.021***	0.022***
Business School	0.103***	0.068***	0.065***
Academic	-0.081***	-0.039***	-0.039***
Honors	0.044***	0.027***	0.026***
Erasmus	0.054***	0.031***	0.029***
Professional transition			
Log(Experience)	0.078***	0.059***	0.059***
Tenure	0.010***	0.009***	0.009**
Unemployment	-0.022***	-0.017***	-0.017***
Family and origin			
Immigrants	0.024***	0.017**	0.012
Excecutive Father	0.032***	0.020***	0.020***
Executive Mother	0.025***	0.017***	0.018***
Localisation			
Paris	0.186***	0.146***	0.153***
IDF	0.134***	0.104***	0.119***
Job characteristics			
Temporary Contrat	-0.090***	-0.066***	-0.064***
Manufacturing	0.029***	0.036***	0.037***
Commercial	-0.049***	-0.001	0.000
Services	-0.057***	-0.034***	-0.033***

** significant to 10%, ** significant to 5%, ***significant to 1%.*
Source: Génération 2004, Census 2006 and DADS 2007.